ADDRESSING BEHAVIORS THAT LEAD TO SHARING FAKE NEWS

Gleb Tsipursky
The Ohio State University

Zachary Morford
Zuce Technologies

Fake news shared by private citizens is sweeping social media. A recent study showed that in the three months before the 2016 US Presidential election, 156 misleading news stories got just under 38 million shares on Facebook (Allcott & Gentzkow, 2017). To address this problem, we propose an intervention, the Pro-Truth Pledge (PTP), which draws on behavioral science research about what causes us to lie and how to prevent such deception, as well as successful strategies in promoting pro-social behaviors from the environmental movement. In taking the pledge, at ProTruthPledge.org, signees agree to abide by twelve behaviors:

- Verify: Fact-check information to confirm it is true before accepting and sharing it.

- Balance: Share the whole truth, even if some aspects do not support my opinion.

- Cite: Share my sources so that others can verify my information.

- Clarify: Distinguish between my opinion and the facts.

- Acknowledge: Acknowledge when others share true information, even when we disagree otherwise

- Reevaluate: Reevaluate if my information is challenged, retract it if I cannot verify it.

- Defend: Defend others when they come under attack for sharing true information, even when we disagree otherwise.

- Align: Align my opinions and my actions with true information.

- Fix: Ask people to retract information that reliable sources have disproved even if they are my allies.

- Educate: Compassionately inform those around me to stop using unreliable sources even if these sources support my opinion.

- Defer: Recognize the opinions of experts as more likely to be accurate when the facts are disputed.

- Celebrate: Celebrate those who retract incorrect statements and update their beliefs toward the truth.

The pledge guidelines encourage those who take the pledge to indicate that they did so by listing themselves in the public database of pledge-takers, posting about taking it on their social media, adding a pledge seal to their website and social media profiles, and noting every time they

share an article on social media that they took the pledge. Taking the pledge serves as a form of public precommitment (see Logue, 1995), where one makes a response now that prevents subsequent impulsive choices. Pre-commitment (or public commitment in this case) is regarded as an effective strategy of behavioral influence (Embry & Biglan, 2008) and self-control, having been shown to positively affect a variety of behaviors including increasing voting (Burgess, Haney, Snyder, Sullivan, & Transue, 2000), increasing recycling (DeLeon & Fuqua, 1995), and both increase participation in retirement savings plans and increase the amount that participants save (Thaler & Benartzi, 2004).

Publicly posting that one has taken the pledge may bring people who sign the pledge into contact with potential social reinforcers (e.g., likes and comments on social media). The pledge seal and other indicators of taking the pledge serve as discriminative stimuli with specific wording such as "I took the Pro-Truth Pledge: please hold me accountable"; sharing that one has taken the pledge may prompt a community evaluative mechanism—other people in one's social media network will be more likely to reinforce social media behavior that accords with the pledge and punish behavior that does not.

The pledge gives people an opportunity to sign up for a newsletter about pledge-related activities and follow its social media, which function as prompts for truth-oriented behaviors. Pledge-takers can also join online and in-person communities dedicated to the pledge. These communities engage in evaluation of various articles, discuss strategies to promote truth-telling in the public sphere, and strategize around spreading the pledge to both private citizens and public figures. Joining and participating in such communities brings people into contact with an environment in which behavior aligning with the pledge is more likely to be reinforced. Research shows that active engagement as a volunteer in promoting such behavior is likely to promote greater engagement in pro-social behavior by the volunteer (Burn, 1991).

Issues the PTP Addresses

The PTP was designed to counteract lying and cognitive biases pertinent to the spread of misinformation. Though cognitive biases come from a cognitive academic tradition, they remain relevant to issues of interest to behavior analysts (e.g., Avila, Yankelevitz, Gonzalez, & Hackenberg, 2013; Macaskill & Hackenberg, 2012, 2013; Wray, Freund, & Dougher, 2009). We can define cognitive biases, generally, as behavioral patterns that deviate from an accepted normative pattern (Kahneman, 2003), and that are considered "errors in judgment." They are considered errors due to their deviation from the normative pattern. Below we consider how the PTP addresses several common cognitive biases.

Confirmation bias refers to our tendency to search for and accept information that aligns with our current beliefs and expectations. One way to address the confirmation bias involves asking people to consider and search for evidence that disproves their initial beliefs (Kray & Galinsky, 2003). The first behavior of the pledge on fact-checking is informed by this research. To ensure clarity on what constitutes such violations, the PTP provides its own definition of misinformation: any information that contradicts verifiable evidence. Misinformation can be provided by lying directly (knowingly providing false or inaccurate information; see Fryling, 2016), lying by omission (when one lies by knowingly leaving out information which alters the interpretation of the information), or misrepresenting the truth to suit one's own purposes. While sometimes misinformation is blatant, other times it is harder to discern. The PTP requires individuals to obtain information from credible fact-checking organizations and the scientific literature. The behaviors

in the pledge that ask people to verify information, cite sources, share the whole truth, reevaluate information, distinguish opinions and facts, and align actions and beliefs with the facts address the confirmation bias.

Two other cognitive biases the PTP addresses are the *in-group bias* and the *Dunning-Kruger effect*. In-group bias refers to when people favor those they perceive to be part of their own group over others not in that group (Verkuyten & Nekuee, 1999). To address the in-group bias, the pledge asks people to (a) defend other people who come under attack when they share accurate information, (b) to request that allies who share inaccurate information retract it, and (c) acknowledge when others share facts even if you otherwise disagree. The Dunning-Kruger effect occurs when those who have less expertise and skills in any given area have an inflated perception of their abilities (Dunning, 2011). To address this problem, the pledge calls on signees to defer to those with expertise.

The PTP also is intended to have an impact on frequency of lying. Studies have shown when people perceive others around them as behaving dishonestly, they also are more likely to behave dishonestly themselves. In turn, if they behave honestly, they perceive others as more likely to behave honestly (e.g., Gino, Norton, & Ariely, 2010). The parallel to seeing one's social media contacts sharing fake news is clear: If one spreads misinformation online through social media, others are likely to do the same, especially if they believe it benefits their in-group (Mazar & Ariely, 2006). By asking pledge-takers to verify and cite their sources, the PTP may make it more likely that others in one's social media network will do the same.

Other research has shown some intriguing findings as to how to increase the likelihood that people will behavior honestly. Reminders about ethical behavior make participants less likely to lie, getting people to sign an honor code or other honesty commitment contract before engaging in tasks in which people were likely to lie increased honesty, and making standards for truthful behavior clear decreased deception (Mazar, Amir, & Ariely, 2008). Our likelihood of lying is strongly impacted by our social network, making it especially important to address social norms around deception (Mann et al. 2014). The last four behaviors of the pledge are specifically intended to encourage pledge-takers to promote honesty in their social networks.

Pro-Truth Pledge Impact: Case Studies

The PTP was launched in March 2017. By January2018 it had over 5000 pledgees, including more than 80 officials and candidates for public office and 300 other public figures, such as academics, journalists, CEOs, and religious and civic leaders. Online and in-person groups dedicated to the PTP have been created in over 20 US states and are continuing to appear in other states as well as in other countries.

We have conducted over 25 informal interviews with pledge-takers to understand their motivations for taking the pledge as well as the impact of the pledge on their behavior. When asked why they take the pledge, people generally report a desire to cast a vote against fake news and demonstrate a personal commitment to honesty. Some also discuss the desire to project a reputation as truth-tellers to enhance their stimulus value as a discriminative or reinforcing stimulus among those with whom they engage; that is, to establish themselves as a discriminative stimulus for providing truthful information.

Several follow-up conversations with pledge-takers suggest that the pledge has impacted behavior. A US Army veteran stated how after taking the pledge, he felt "an open commitment to a certain attitude" to "think hard when I want to play an article or statistic which I'm not completely

sold on." He found the pledge had powerful stimulus control effects as it "really does seem to change one's habits," helping push him both to correct his own mistakes with an "attitude of humility and skepticism, and of honesty and moral sincerity," and also to encourage "friends and peers to do so as well." Similarly, a Christian pastor and community leader described how he "took the Pro-Truth Pledge because I expect our political leaders at every level of government to speak truth and not deliberately spread misinformation to the people they have been elected to serve. Having taken the pledge myself, I put forth the effort to continually gather information validating stories and headlines before sharing them on my social media outlets."

We also have observed instances where the pledge was involved with people retracting statements. For instance, a candidate for Congress in Idaho posted on his Facebook wall a screenshot of a tweet by Donald Trump criticizing minority and disabled children. After being called out on it by fellow pledge-takers, he went and searched Trump's feed. He could not find the original tweet, and while Trump may have deleted that tweet, the candidate edited his own Facebook post to say that "Due to a Truth Pledge I have taken I have to say I have not been able to verify this post." He indicated that he would be more careful with future postings. In another case a photographer from New York shared and endorsed an article from *OccupyDemocrats.com*, a site shown by credible fact-checkers used by the PTP to be systematically unreliable. Other pledge-takers asked him to withdraw it, and he did so.

We are currently undertaking a more formal and quantitative evaluation of the impact of the pledge, with a focus on whether sharing of information on social media is impacted by the pledge. We decided to target Facebook, as the most popular social media platform, to compare people's sharing on Facebook before and after taking the pledge. While not a controlled study, this provides tentative preliminary evidence as to the effects of the PTP on sharing truthful information.

Conclusion

The PTP uses behavioral science research to address fake news sharing on social media, and evidence clearly shows that people are willing to take the pledge. Regarding its effectiveness in behavior change, case study evidence from both private citizens and public figures show a self-reported impact on behavior and instances of external observed changes in behavior. Further evidence in the form of longitudinal as well as controlled research needs to be done to test the empirical effect of this strategy. However, even at this stage we comfortably recommend that strong efforts should be made to encourage private citizens and public figures alike to go to ProTruthPledge.org and sign the pledge to address the kind of behaviors that lead to the sharing of fake news.

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211-36. doi: 10.1257/jep.31.2.211

Avila, R., Yankelevitz, R. L., Gonzalez, J. C., & Hackenberg, T. D. (2013). Varying the costs of sunk costs: Optimal and non-optimal choices in a sunk-cost task with humans. *Journal of the experimental analysis of behavior*, *100*(2), 165-173. doi: 10.1002/jeab.42

Burn, S. M. (1991). Social psychology and the stimulation of recycling behaviors: The block leader approach. *Journal of applied social psychology*, *21*(8), 611-629. doi: 10.1111/j.1559-1816.1991.tb00539.x

Burgess, D., Haney, B., Snyder, M., Sullivan, J. L., & Transue, J. E. (2000). Rocking the vote: Using personalized messages to motivate voting among young adults. *Public Opinion Quarterly*, *64*(1), 29-52. doi: 10.1086/316758

De Leon, I. G., & Fuqua, R. W. (1995). The effects of public commitment and group feedback on curbside recycling. *Environment and behavior*, *27*(2), 233-250. doi: 10.1177/0013916595272007

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology* (Vol. 44, pp. 247-296). Academic Press. doi: 10.1016/B978-0-12-385522-0.00005-6

Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical child and family psychology review*, *11*(3), 75-113. doi: 10.1007/s10567-008-0036-x

Fryling, M. J. (2016). A Developmental-Behavioral Analysis of Lying. *International Journal of Psychology and Psychological Therapy*, *16*(1).

Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological science*, *21*(5), 712-720. doi: 10.1177/0956797610366545

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, *69*(6), 1069.

Logue, A. W. (1995). *Self-control: Waiting until tomorrow for what you want today*. Prentice-Hall, Inc.

Macaskill, A. C., & Hackenberg, T. D. (2012). Providing a reinforcement history that reduces the sunk cost effect. *Behavioural Processes*, *89*(3), 212-218. doi: 10.1016/j.beproc.2011.11.001

Macaskill, A. C., & Hackenberg, T. D. (2013). Optimal and nonoptimal choice in a laboratory-based sunk cost task with humans: A cross-species replication. *Journal of the experimental analysis of behavior*, *100*(3), 301-315. doi**:** 10.1002/jeab.52

Mann, H., Garcia-Rada, X., Houser, D., & Ariely, D. (2014). Everybody else is doing it: Exploring social transmission of lying behavior. *PloS one*, *9*(10), e109591. doi: 10.1371/journal.pone.0109591

Mazar, N., Amir, O., & Ariely, D. (2008). More ways to cheat-expanding the scope of dishonesty. *Journal of Marketing Research*, *45*(6), 651-653.

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of public policy & Marketing*, *25*(1), 117-126. doi: 10.1509/jppm.25.1.117

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, *112*(S1), S164-S187. doi: 10.1086/380085

Verkuyten, M., & Nekuee, S. (1999). Ingroup bias: the effect of self-stereotyping, identification and group threat. *European Journal of Social Psychology*, *29*(2-3), 411-418. doi: 10.1002/(SICI)1099-0992(199903/05)29:2/3<411::AID-EJSP952>3.0.CO;2-8

Wray, A. M., Freund, R. A., & Dougher, M. J. (2009). A behavior-analytic account of cognitive bias in clinical populations. *The Behavior Analyst*, *32*(1), 29-49. doi: 10.1007/BF03392174